URL: http://www.elsevier.nl/locate/entcs/volume84.html 12 pages

Statistics of implicational logic [1]

Marek Zaionc [2]

Computer Science Department
Jagiellonian University
Krakow, Poland

Abstract

In this paper we investigate the size of the fraction of tautologies of the given length n against the number of all formulas of length n for implicational logic. We are specially interested in asymptotic behavior of this fraction. We demonstrate the relation between a number of premises of implicational formula and asymptotic probability of finding formula with this number of premises. Furthermore we investigate the distribution of this asymptotic probabilities. Distribution for all formulas is contrasted with the same distribution for tautologies only. We prove those distributions to be so different that enable us to estimate likelihood of truth for a given long formula. Despite of the fact that all discussed problems and methods in this paper are solved by mathematical means, the paper may have some philosophical impact on the understanding how much the phenomenon of truth is sporadic or frequent in random logical sentences.

1 Introduction

Probabilistic methods appear to be very powerful in combinatorics and computer science. A point of view of those methods is that we investigate the typical object chosen from the set. In this paper we investigate the proportion between the number of tautologies of the given length n against the number of all formulas of length n for propositional formulas. Our interest lays in finding limit of that fraction when $n \to \infty$. If the limit exists it represents the real number between 0 and 1 which we may call *the density of truth* for the logic investigated. In general we are interested in finding the *"density"* of some other classes of formulas. This paper is a part of the ongoing research in which we try to estimate the likelihood of truth for the given propositional

[1] Supported by the State Committee for Scientific Research in Poland (KBN), research grant 7T11C 022 21
[2] Email: zaionc@ii.uj.edu.pl

logic. Consult for example paper [3] for purely implicational logic of one variable (and at the same time a type system) and [7] for the classical logic of implication and negation. In the paper [8] we have found the exact proportion between intuitionistic and classical logics of the same language.

In this paper we investigate the language $\mathcal{F}_k^{\{\to\}}$ consisting of implicational formulas over k propositional variables.

Definition 1.1 *The language $\mathcal{F}_k^{\{\to\}}$ (over k propositional variables) consists of propositional k variables $\{a_1, \ldots, a_k\}$ and it is closed by implication $\to$, i.e.*

$$a_i \in \mathcal{F}_k^{\{\to\}} \quad \forall i \leq k$$
$$\phi \to \psi \in \mathcal{F}_k^{\{\to\}} \ \textit{if} \ \phi \in \mathcal{F}_k^{\{\to\}} \ \textit{and} \ \psi \in \mathcal{F}_k^{\{\to\}}$$

First we have to establish the way the length of formulas are measured.

Definition 1.2 *By $\|\phi\|$ we mean the length of formula ϕ which we define as the total number of occurrences of propositional variables in the formula. Parenthesis which are sometimes necessary and implication sign itself are not included in the length of formula. Formally, $\|a_i\| = 1$ and $\|\phi \to \psi\| = \|\phi\| + \|\psi\|$.*

Definition 1.3 *We associate the density $\mu(\mathcal{A})$ with a subset $\mathcal{A} \subset \mathcal{F}_k^{\{\to\}}$ of formulas as:*

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{\#\{t \in \mathcal{A} : \|t\| = n\}}{\#\{t \in \mathcal{F}_k^{\{\to\}} : \|t\| = n\}} \tag{1}$$

if the limit exists.

The number $\mu(\mathcal{A})$ if exists is an asymptotic probability of finding formula from the class $\mathcal{A}$ among all formulas from $\mathcal{F}_k^{\{\to\}}$ or it can be interpreted as the asymptotic density of the set $\mathcal{A}$ in the set $\mathcal{F}_k^{\{\to\}}$. It can be seen immediately that the density μ is finitely additive so if $\mathcal{A}$ and $\mathcal{B}$ are disjoint classes of formulas such that $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ exist then $\mu(\mathcal{A} \cup \mathcal{B})$ also exists and $\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B})$. It is straightforward to observe that for any finite set $\mathcal{A}$ the density $\mu(\mathcal{A})$ exists and is 0. Dually for co-finite sets $\mathcal{A}$ the density $\mu(\mathcal{A}) = 1$. The density μ is not countably additive so in general the formula

$$\mu\left(\bigcup_{i=0}^{\infty} \mathcal{A}_i\right) = \sum_{i=0}^{\infty} \mu\left(\mathcal{A}_i\right) \tag{2}$$

it is not true for all pairwise disjoint classes of sets $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$. The good counterexample for the equation (2) is to take as $\mathcal{A}_i$ the singleton consisting of i-th formula from our language. On the left hand side of (2) we get $\mu\left(\mathcal{F}_k^{\{\to\}}\right)$ which is 1 but on right hand side $\mu\left(\mathcal{A}_i\right) = 0$ for all $i \in \mathbb{N}$.
In the paper we are specially interested in distribution of densities with respect of some numerical property of formulas.

Definition 1.4 *By a random variable* X *we understand the function* $X : \mathcal{F}_k^{\{\to\}} \mapsto \mathbb{N}$ *which assigns a number* $n \in \mathbb{N}$ *to the implicational formula in such a way that for any* n *the density* $\mu\left(\left\{\phi \in \mathcal{F}_k^{\{\to\}} : X(\phi) = n\right\}\right)$ *exists and moreover*

$$\sum_{n=0}^{\infty} \mu\left(\left\{\phi \in \mathcal{F}_k^{\{\to\}} : X(\phi) = n\right\}\right) = 1.$$

Definition 1.5 *By the distribution of random variable* X *we mean the function* $\overline{X} : \mathbb{N} \mapsto \mathbb{R}$ *defined by:*

$$\overline{X} \ : \ \mathbb{N} \ni n \mapsto \mu\left(\left\{\phi \in \mathcal{F}_k^{\{\to\}} : X(\phi) = n\right\}\right) \in \mathbb{R}$$

Definition 1.6 *The expected value, variance and standard deviation are defined in conventional way by:* $E(\overline{X}) = \sum_{p=0}^{\infty} p \cdot \overline{X}(p)$ *and* $D^2(\overline{X}) = E(\overline{X}^2) - (E(\overline{X}))^2 = \sum_{p=0}^{\infty} p^2 \overline{X}(p) - (E(\overline{X}))^2$ *so the standard deviation of* X *is* $\sqrt{\mathcal{D}^2(\overline{X})}$

In paper [3] we showed what is the relation between the number of premises of implicational formula and asymptotic probability of finding formula with this number of premises. In this paper we are going to investigate the distribution of densities with respect of the number of premises but only for simple tautologies, which form a huge subset of all tautologies. We prove that this distribution is so different from the previous one that it can be used to distinguish a tautology only by counting the number of its premises.

2 Counting Formulas

In this section we present some properties of numbers characterizing the amount of formulas in different classes defined in our language. For general technique for construction of combinatorial coefficients, called analytic combinatorics, used in theorem and lemmas 2.2, 2.5 and 2.7 consult an excellent presentation in [2].

Definition 2.1 *By* F_n^k *we mean the total number of formulas from* $\mathcal{F}_k^{\{\to\}}$ *of the length* n *so:* $F_n^k = \#\{\phi \in \mathcal{F}_k^{\{\to\}} : \|\phi\| = n\}$.

Lemma 2.2 *Numbers* F_n^k *are given by the following recursion on* n*:* $F_0^k = 0$, $F_1^k = k$ *and* $F_n^k = \sum_{i=1}^{n-1} F_i^k F_{n-i}^k$.

Proof. For $n = 0$ and $n = 1$ it is obvious. Any formula of length $n > 1$ is the implication between some pair of formulas of lengths i and $n - i$, respectively. Therefore the total number of such pairs is $\sum_{i=1}^{n-1} F_i^k F_{n-i}^k$. □

Lemma 2.3 *Let* C_n *be* n*-th Catalan number. Then the number* $F_n^k = k^n C_n$.

Proof. Proof is straightforward. □

For more elaborate treatment of Catalan numbers see for example [6, pp. 43–44].

Definition 2.4 *By $F_n^k(p)$ we mean the number of formulas of length n having p premises, i.e. formulas which are of the form: $\tau = \tau_1 \to (\ldots \to (\tau_p \to \alpha))$, where α is a propositional variable.*

Since numbers $F_n^k(p)$ are the cardinalities of disjoint sets of formulas for different p's and since there are no formulas of length n having more then $n-1$ premises, for $n \geq 2$ we have $F_n^k = F_n^k(1) + \ldots + F_n^k(n-1)$. Consequently by $C_n(p)$ we mean $F_n^1(p)$. As in lemma 2.3 we have

$$F_n^k(p) = k^n C_n(p). \tag{3}$$

Lemma 2.5 *The number $F_n^k(p)$ is given by the following recursion on p:*

$$\begin{aligned}
F_n^k(0) &= \textit{if } n = 1 \textit{ then } k \textit{ else } 0, \\
F_n^k(1) &= \textit{if } n = 0 \textit{ then } 0 \textit{ else } F_{n-1}^k, \\
F_n^k(p) &= \sum_{i=1}^{n-p} F_i^k F_{n-i}^k(p-1).
\end{aligned}$$

Proof. The proof is technical. □

We are going to isolate the class of simple tautologies which is an important and huge fragment of the set of tautologies. As we will see afterwards the class of simple tautologies is big enough to be a good approximation of the whole set of tautologies. Therefore investigations about behavior of the whole set can be estimated by this fragment.

Definition 2.6 *A simple tautology is a formula $\tau \in \mathcal{F}_k^{\{\to\}}$ on the form $\tau = \tau_1 \to (\ldots \to (\tau_n \to \alpha)\ldots)$ such that there is at least one component τ_i identical with α.*

Evidently, a simple tautology is a tautology. Let G_n^k be the number of simple tautologies of length n built with k propositional variables and $G_n^k(p)$ be the number of simple tautologies of length n built with k variables with p premises. Our goal is to find how big asymptotically is the fragment of simple tautologies within the set of all formulas.

Lemma 2.7 *The number G_n^k of simple tautologies is given by the recursion $G_1^k = 0$, $G_2^k = k$ and $G_n^k = \sum_{i=2}^{n-1} F_{n-i}^k G_i^k + (F_{n-1}^k - G_{n-1}^k)$.*

Proof. The proof is based on two observations: First, $\tau_1 \to \tau_2$ is simple tautology if τ_2 is. So for every formula τ_1 of length $n-i$ and every simple tautology τ_2 of length i we have one simple tautology $\tau_1 \to \tau_2$ of length n. The sum starts from $i = 2$ because there are no simple tautologies of length 1. This part is responsible for the component $\sum_{i=2}^{n-1} F_{n-i}^k G_i^k$. The only other simple tautologies are those for which τ_1 is a propositional variable identical with the propositional variable the formula τ_2 points to. □

Lemma 2.8 *The number $G_n^k(p)$ of simple tautologies with p premises is given by the following recursion on p,*

$$G_n^k(0) = \textit{if } n = 1 \textit{ then } k \textit{ else } 0,$$

$$G_n^k(p+1) = \begin{cases} 0 & \textit{if } n \le p \\ \sum_{i=2}^{n-1} F_{n-i}^k G_i^k(p) + (F_{n-1}^k(p) - G_{n-1}^k(p)) & \textit{if } n > p \end{cases}$$

Proof. The same argument as in lemma 2.7. Proof must be accompanied with counting the number of premises of the considered simple tautology. □

3 Generating functions

The main tool we use for dealing with asymptotics of sequences of numbers are *generating functions.* A nice exposition of the method can be found in [6] and [1]. See also the recent works on random combinatorial structures in [2]. For the presentation of this method in logics consult also papers [7], [8] and [3]. Many questions concerning the asymptotic behavior of a sequence A can be efficiently resolved by analyzing the behavior of generating function f_A at the complex circle $|z| = R$. The key tool will be the following result due to Szegö [5] [Thm. 8.4], see as well [6] [Thm. 5.3.2] which relates the generating functions of numerical sequences with limit of fractions. For the technique of proof described below please consult also [3]. We need the following much simpler version of the Szegö Lemma.

Lemma 3.1 *(simplified Szegö lemma) Let $v(z)$ be analytic in $|z| < 1$ with $z = 1$ the only singularity at the circle $|z| = 1$. If $v(z)$ in the vicinity of $z = 1$ has an expansion of the form $v(z) = \sum_{p \ge 0} v_p (1-z)^{\frac{p}{2}}$, where $p > 0$, and the branch chosen above for the expansion equals to $v(0)$ for $z = 0$, then $[z^n]\{v(z)\} = v_1 \binom{1/2}{n}(-1)^n + O(n^{-2})$.*

The symbol $[z^n]\{v(z)\}$ stands for the coefficient of z^n in the exponential series expansion of $v(z)$. For technical reasons we will need to know the rate of growth of the function $\binom{1/2}{n}(-1)^n$ which appears at the formula (3.1)

Lemma 3.2 *For $n \in \mathbb{N}$ we have*

$$\binom{1/2}{n}(-1)^{n+1} = O(n^{-3/2}) \tag{4}$$

Proof. It can be obtained by the Stirling approximation formula

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \tag{5}$$

(see [4] for details, consult also lemma 7.5 page 589 at [3]). □

In this part of the section we are going to present the method of finding asymptotic densities for the classes of formulas for which the generating functions are already calculated. The main tool used for this purpose is theorem based on simplified Szegö lemma. The following lemma is a main tool for finding limits of the fraction $\frac{a_n}{b_n}$, when generating functions for sequences a_n and b_n satisfies conditions of simplified Szegö lemma 3.1.

Lemma 3.3 *Suppose two functions $v(z)$ and $w(z)$ satisfy assumptions of simplified Szegö lemma (3.1) i.e. both v and w are analytic in $|z| < 1$ with $z = 1$ being the only singularity at the circle $|z| = 1$. Both $v(z)$ and $w(z)$ in the vicinity of $z = 1$ have expansions of the form $v(z) = \sum_{p \geq 0} v_p (1-z)^{p/2}$, and $w(z) = \sum_{p \geq 0} w_p (1-z)^{p/2}$, then the limit of $\frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}}$ exists and is given by formula:*

$$\lim_{n \to \infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{v_1}{w_1} \tag{6}$$

Proof. Applying the main formula from simplified Szegö lemma 3.1 and equation (4) from lemma 3.2 we obtain

$$\lim_{n \to \infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \lim_{n \to \infty} \frac{v_1 \binom{1/2}{n} (-1)^n + O(n^{-2})}{w_1 \binom{1/2}{n} (-1)^n + O(n^{-2})} = \frac{v_1}{w_1} \tag{7}$$

□

Theorem 3.4 *Suppose two functions $v(z)$ and $w(z)$ satisfy assumptions of simplified Szegö lemma (lemma 3.1). Suppose we have functions $\widetilde{v}$ and $\widetilde{w}$ satisfying $\widetilde{v}(\sqrt{1-z}) = v(z)$ and $\widetilde{w}(\sqrt{1-z}) = w(z)$ then the limit of $\frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}}$ exists and is given by formula: $\lim_{n \to \infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{(\widetilde{v})'(0)}{(\widetilde{w})'(0)}$.*

Proof. Simple consequence of lemma 3.3. New functions $\widetilde{v}$ and $\widetilde{w}$ have expansions $\widetilde{v}(z) = \sum_{p \geq 0} v_p z^p$, and $\widetilde{w}(z) = \sum_{p \geq 0} w_p z^p$. Therefore $v_1 = (\widetilde{v})'(0)$ and $w_1 = (\widetilde{w})'(0)$. □

4 Calculating generating functions

We start with calculating generating functions for all recursively defined sequences from the section 2.

Lemma 4.1 *The generating function f_F for the numbers F_n^k is $f_F(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\,k\,z}$,*

Proof. Straightforward. See for example [3]. □

As a special case, when $k = 1$ the generating function f_C for Catalan numbers is given by $f_C(z) = 1/2 - (\sqrt{1-4z})/2$.

Lemma 4.2 *For fixed $p \geq 0$ the generating functions $f_{C(p)}$ and $f_{F(p)}$ for $C_n(p)$ and $F_n^k(p)$, respectively are the following:*

$$f_{C(p)}(z) = z \cdot (f_C(z))^p \;=\; z \cdot \left(\frac{1 - \sqrt{1-4z}}{2}\right)^p, \tag{8}$$

$$f_{F(p)}(z) = k \cdot z \cdot (f_F(z))^p \;=\; k \cdot z \cdot \left(\frac{1 - \sqrt{1-4kz}}{2}\right)^p. \tag{9}$$

Proof. Let $f_{C(p)}(z)$ be a generating function for $C_n(p)$. Lemma 2.5 gives $C_n(p) = \sum_{i=1}^{n-p} C_i C_{n-i}(p-1)$ which becomes after a closer examination the equality $f_{C(p-1)}(z) \cdot f_C(z) = f_{C(p)}(z)$. Since $C_n(1) = C_{n-1}$ we get $f_{C(1)}(z) = z(f_C(z))$ and consequently $f_{C(p)}(z) = z(f_C(z))^p$. Thanks to equation (3) we get $f_{F(p)}(z) = f_{C(p)}(kz)$ which ends the proof of (9.) Notice that formulas (8) and (9) are also correct for $p = 0$. $\square$

Lemma 4.3 *The generating function f_G for numbers G_n^k is:*

$$f_G(z) = \frac{z f_F(z)}{1 - f_F(z) + z} = -\frac{-\frac{1}{2}z + \frac{1}{2}\, z\sqrt{1-4\,k\,z}}{\frac{1}{2} + \frac{1}{2}\sqrt{1-4\,k\,z} + z}. \tag{10}$$

Proof. The recurrence given by equation from lemma 2.7 becomes $f_G = f_G \cdot f_F + z \cdot f_F - z \cdot f_G$. Solving it gives the solution above. $\square$

Lemma 4.4 *For fixed p the generating function $f_{G(p)}$ for $G_n^k(p)$ can be defined by the following recursion on p: $f_{G(0)}(z) = 0$ and*

$$f_{G(p+1)}(z) = f_F(z) \cdot f_{G(p)}(z) + kz^2\,(f_F(z))^p - z f_{G(p)}(z). \tag{11}$$

Proof. Formula for $f_{G(p+1)}$ is a simple encoding of the recurrence (4). Multiplication $f_F(z) \cdot f_{G(p)}(z)$ is responsible for the fragment $\sum_{i=2}^{n-1} F_{n-i}^k G_i^k(p)$. According to formula (9) (see lemma 4.2) for functions $F_n^k(p)$ we have that $kz\,(f_F(z))^p$ stands for $F_n^k(p)$. Since the number in recurrence depends on $n-1$ not on n it have to be additionally multiply by z. The last fragment $z f_{G(p)}(z)$ is responsible for the recursion $G_{n-1}^k(p)$ in (4). $\square$

5 Calculation of limits

In this section we are going to find asymptotic densities for the classes of formulas for which the generating functions are already calculated. The main tool used for this purpose is a theorem 3.4.

First we recall two results from [3]. In the first one we consider the probability that the given formula is a simple tautology. The meaning of this theorem is that the limit of the fraction G_n^k/F_n^k while n tends to infinity exists and the size of the set of all tautology formulas is at least as big as $O(1/k)$. The second theorem finds the probability for the given formula to has p premises. Both are good examples of usefulness of theorem 3.4.

Theorem 5.1 *The asymptotic probability of the fact that a random formula is a simple tautology is:*

$$\lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \frac{4k+1}{(2k+1)^2} \tag{12}$$

Proof. Indeed, first we recall equation (10) from lemma 4.3 for f_G and formula from lemma 4.1 for f_F. In order to satisfy assumptions of theorem 3.4 let us normalize functions f_G and f_F in such a way to have only singularity located in $|z| \leq 1$ at the position in $z = 1$. So, let us define functions $\overline{f_G}(z) =$

$f_G(z/(4k))$ and $\overline{f_F}(z) = f_F(z/(4k))$. After cancellation we get: $\overline{f_G}(z) = -k+\frac{1}{2}-\frac{1}{2}\sqrt{1-z}+2\frac{k^2}{-\sqrt{1-z}+2k+1}$ and $\overline{f_F}(z) = \frac{1}{2}-\frac{1}{2}\sqrt{1-z}$. This representation reveals that the only singularity of $\overline{f_G}(z)$ located in $|z| \leq 1$ is indeed $z = 1$. The other singularity, located where the denominator in the last fraction becomes 0, is achieved when $\sqrt{1-z} = 2k+1$, i.e., for z of a modulus substantially greater than 1. For function $\overline{f_F}$ it is clear. We have to remember that a change of a caliber of the radius of convergence for functions f_G and f_F effects accordingly sequences represented by the new functions. Therefore we get $G_n^k = (4k)^n \left([z^n]\{\overline{f_G}(z)\}\right)$ and $F_n^k = (4k)^n \left([z^n]\{\overline{f_F}(z)\}\right)$. Now we are ready to use theorem 3.4. Let us define functions $\widetilde{f_F}$ and $\widetilde{f_G}$ so as to satisfy equalities: $\widetilde{f_F}(\sqrt{1-z}) = \overline{f_F}(z)$ and $\widetilde{f_G}(\sqrt{1-z}) = \overline{f_G}(z)$. Functions $\widetilde{f_F}$ and $\widetilde{f_G}$ are: $\widetilde{f_G}(z) = -\frac{1}{2}\frac{(1-z^2)(z-1)}{2k+2kz+1-z^2}$ and $\widetilde{f_F}(z) = \frac{1}{2}-\frac{1}{2}z$. Derivatives of $\widetilde{f_F}$ and $\widetilde{f_G}$ are the following: $(\widetilde{f_G})'(z) = -\frac{1}{2}\frac{-(1+4k-z)(z-1)}{(z-2k-1)^2}$ and $(\widetilde{f_F})'(z) = -\frac{1}{2}$. Finally derivatives $(\widetilde{f_G})'(0) = -\frac{1}{2}\frac{4k+1}{(2k+1)^2}$ and $(\widetilde{f_F})'(0) = -\frac{1}{2}$. Now applying theorem 3.4 we get $\lim_{n\to\infty}\frac{G_n^k}{F_n^k} = \lim_{n\to\infty}\frac{(4k)^n\left([z^n]\{\overline{f_G}(z)\}\right)}{(4k)^n\left([z^n]\{\overline{f_F}(z)\}\right)} = \frac{(\widetilde{f_G})'(0)}{(\widetilde{f_F})'(0)} = \frac{4k+1}{(2k+1)^2}$ which ends the proof. □

The proof of the theorem 5.1 reveals the technique of showing the convergence of fractions in which both nominator and denominator are given recursively and both generating functions satisfy Szegö lemma. The proof of the next theorem will use exactly the same method.

Theorem 5.2 *The asymptotic probability of the fact that a random formula admits exactly p premises is:*

$$\lim_{n\to\infty}\frac{F_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}} \tag{13}$$

Proof. First we recall equation (9) from lemma 4.2 describing function $f_{F(p)}$. All steps for denominator f_F are already done in the previous theorem. Function $\overline{f_{F(p)}}(z) = f_{F(p)}(z/(4k))$ defined to satisfy theorem 3.4 is as follows: $\overline{f_{F(p)}}(z) = \frac{z}{4}(\overline{f_F(z)})^p = \frac{z}{4}\left(\frac{1-\sqrt{1-z}}{2}\right)^p$. It is clear that $\overline{f_{F(p)}}(z)$ admits the only singularity at $z = 1$. As in previous theorem let us define function $\widetilde{f_{F(p)}}$ so as to satisfy the following equations: $\widetilde{f_{F(p)}}(\sqrt{1-z}) = \overline{f_F(p)}(z)$. Therefore $\widetilde{f_{F(p)}}(z) = \frac{1-z^2}{4}\left(\frac{1-z}{2}\right)^p$. Derivative of the function $\widetilde{f_{F(p)}}(z)$ is: $(\widetilde{f_{F(p)}})'(z) = -\frac{z^2}{2}\left(\frac{1-z}{2}\right)^2 - p\frac{(1-z^2)}{8}\left(\frac{1-z}{2}\right)^{p-1}$ so $(\widetilde{f_{F(p)}})'(0) = -\frac{1}{2}\frac{p}{2^{p+1}}$ which concludes the proof. □

The main goal of this section is to find the term for the asymptotic density of classes of simple tautologies with p premises which allows us to speak about distribution of probabilities.

Theorem 5.3 *The asymptotic probability of the fact that a random formula is a simple tautology with exactly p premises is:*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}. \tag{14}$$

Proof. Base consequently on lemma 4.4 and the main theorem 3.4. Generating functions in lemma 4.4 are defined recursively on p. Using the same technique as in two previous theorems we will find, also recursive on p, terms for appropriate limits. All steps for denominator f_F are already done in theorem 5.1. Functions $\overline{f_{G(p)}}(z) = f_{G(p)}(z/(4k))$ are defined to satisfy theorem 3.4 and are as follows: $\overline{f_{G(0)}}(z) = 0$ and

$$\overline{f_{G(p+1)}}(z) = \overline{f_F}(z) \cdot \overline{f_{G(p)}}(z) + \frac{z^2}{16k}\left(\overline{f_F}(z)\right)^p - \frac{z}{4k}\overline{f_{G(p)}}(z). \tag{15}$$

It is clear that for every p function $\overline{f_{G(p)}}(z)$ admits the only singularity at $z = 1$. As in previous theorem let us define function $\widetilde{f_{G(p)}}$ so as to satisfy the following equation: $\widetilde{f_{G(p)}}(\sqrt{1-z}) = \overline{f_G(p)}(z)$. Therefore $\widetilde{f_{G(0)}}(z) = 0$ and

$$\widetilde{f_{G(p+1)}}(z) = \left(\widetilde{f_F}(z) - \frac{1-z^2}{4k}\right) \cdot \widetilde{f_{G(p)}}(z) + \frac{(1-z^2)^2}{16k}\left(\widetilde{f_F}(z)\right)^p. \tag{16}$$

Now we define two sequences of real numbers $h_p = \widetilde{f_{G(p)}}(0)$ and $g_p = (\widetilde{f_{G(p)}})'(0)$. Remember that $\widetilde{f_F}(0) = \frac{1}{2}$ and $(\widetilde{f_F})'(0) = -\frac{1}{2}$ (see proof of theorem 5.1). Therefore sequences h_p and g_p are given by $h_0 = 0$ and $h_{p+1} = \left(\frac{2k-1}{4k}\right) \cdot h_p + \frac{1}{k2^{p+4}}$ and $g_0 = 0$ and $g_{p+1} = \left(\frac{2k-1}{4k}\right) \cdot g_p - \frac{1}{2} \cdot h_p + \frac{1}{k2^{p+4}}$. Solving the system of mutual recursion using the standard generating functions technique we get $h_p = \frac{1}{2^{p+2}} - \frac{1}{4}\left(\frac{2k-1}{4k}\right)^p$ and consequently $g_p = -\frac{1}{2}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right)$ which concludes the proof. □

Theorem 5.4 *The asymptotic probability of the fact that a random simple tautology has exactly p premises is:*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{G_n^k} = \frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right). \tag{17}$$

Proof. Combine two limit equations from theorems 5.1 and 5.3. □

6 Distribution of probabilities

In this section we will discuss and compare the distribution of probabilities proved in previous sections. There are two main questions we wish to discuss:

What is a probability that randomly chosen implicational formula admits p ***premises ?***

What is a probability that randomly chosen implicational simple tautology admits p ***premises ?***

To answer the first question we group together all formulas with p premises and according to the definition (1) we try to find the asymptotic probability of this class. But this is exactly what we have found in theorem 5.2. So let us start with analyzing the first distribution:

Definition 6.1 *Let us define the random variable X which assigns to an implicational formula the number of its premises.*

Theorem 6.2 *Random variable X has the distribution: $\overline{X}(p) = \frac{p}{2^{p+1}}$. Expected value $E(\overline{X}) = 3$, variance $D^2(\overline{X}) = 4$. The standard deviation of X is 2.*

Proof. As we know the number of formulas of length n with the p premises is $F_n^k(p)$. Therefore, according to theorem 5.2, the asymptotic probability is $\lim_{n\to\infty} \frac{F_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}}$ (see theorem 5.2). This form a distribution since $\sum_{p=0}^{\infty} \frac{p}{2^{p+1}} = 1$. Expected value $E(\overline{X}) = \sum_{p=1}^{\infty} p\,\overline{X}(p) = \sum_{p=1}^{\infty} p \frac{p}{2^{p+1}} = 3$ and variance $D^2(\overline{X}) = E(\overline{X}^2) - (E(\overline{X}))^2 = \sum_{p=1}^{\infty} p^2 \frac{p}{2^{p+1}} - 9 = 4$, so the standard deviation of X is $\sqrt{\mathcal{D}^2(\overline{X})} = 2$. □

¿From the whole discussion we can see that surprisingly, a typical implicational formula is supposed to have exactly 3 premisses. For example the amount of formulas with number of premises between 1 and 5 ie. which are typical ± standard deviation is 57/64 which is about 89%.
Now we will answer the second question. First we have to isolate the class of all simple tautologies with p premises and compare it to the class of all simple tautologies. But this is exactly what we have found in theorem 5.4. We will see now the substantial difference between distribution of the number of premises for all formulas and the same distribution for simple tautologies only.

Definition 6.3 *For every $k \geq 1$ let us separately define the random variable Y_k which assigns to an implicational simple tautology in the language $\mathcal{F}_k^{\{\to\}}$ the number of its premises.*

Theorem 6.4 *Random variable Y_k has the following distribution, expected value and variance: $\overline{Y_k}(p) = \frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right)$, $E(\overline{Y_k}) = \frac{40k^2+18k+3}{(2k+1)(4k+1)}$ and $D^2(\overline{Y_k}) = \frac{384k^4+288k^3+160k^2+48k+4}{(2k+1)^2(4k+1)^2}$*

Proof. Trivial technical calculations are omitted. Nevertheless notice that $\lim_{k\to\infty} E(\overline{Y_k}) = 5$ and $\lim_{k\to\infty} D^2(\overline{Y_k}) = 6$. □

7 Limit distribution

The natural question is how the distribution of true sentences looks like for very large numbers k, or is there a uniform asymptotic distribution when k, the number of propositional variables in the logic, tends to infinity. The answers are following:

Lemma 7.1 *For fixed number of premises $p \geq 0$*

(18)
$$\lim_{k\to\infty} \frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right) = \frac{p(p-1)}{2^{p+2}}.$$

Proof. For $p = 0$ and $p = 1$ it is obvious. For $p \geq 2$ it is a simple limit exercise. $\square$

Let us name the limit distribution by $\overline{Y_\infty}(p) = \frac{p(p-1)}{2^{p+2}}$. This is in fact distribution since $\sum_{p=0}^{\infty} \overline{Y_\infty}(p) = \sum_{p=0}^{\infty} \frac{p(p-1)}{2^{p+2}} = 1$. Expected value of $\overline{Y_\infty}$ is $E(\overline{Y_\infty}) = \sum_{p=0}^{\infty} p\,\overline{Y_\infty}(p) = \sum_{p=0}^{\infty} \frac{p^2(p-1)}{2^{p+2}} = 5$. The variance of $\overline{Y_\infty}$ is $D^2(\overline{Y_\infty}) = E((\overline{Y_\infty} - E(\overline{Y_\infty}))^2) = E((\overline{Y_\infty})^2) - (E(\overline{Y_\infty}))^2 = \sum_{p=0}^{\infty} p^2 \frac{p(p-1)}{2^{p+2}} - 25 = 31 - 25 = 6$. Now it is clear that for every $p \geq 0$ $\lim_{k\to\infty} \overline{Y_k}(p) = \overline{Y_\infty}(p)$ and $\lim_{k\to\infty} E(\overline{Y_k}) = E(\overline{Y_\infty})$ and also $\lim_{k\to\infty} D^2(\overline{Y_k}) = D^2(\overline{Y_\infty})$ (see the proof of theorem 6.4).
The componentwise convergence (with respect to p) presented in lemma 7.1 and summarized by the formula $\lim_{k\to\infty} \overline{Y_k}(p) = \overline{Y_\infty}(p)$ can be extended to much stronger uniform convergence. Bellow we show the uniformity of convergence of the sequence of distributions $\overline{Y_k}$ to $\overline{Y_\infty}$ when k tends to infinity. Therefore the distribution $\overline{Y_\infty}$ can be treated as a good model of distribution for simple tautologies for the language $\mathcal{F}_k^{\{\to\}}$ when the number k of atomic propositional variables is large.

Theorem 7.2 *The sequence of distributions $\overline{Y_k}$ uniformly converges to the distribution $\overline{Y_\infty}$.*

Proof. It can be shown by very laborious but simple calculations of the cartesian distance between distributions $\overline{Y_k}$ and $\overline{Y_\infty}$. The distance between functions is defined by $dis(f,g) = \sum_{p=0}^{\infty} (f(p) - g(p))^2$. Since we have explicit formulas for $\overline{Y_k}$ and $\overline{Y_\infty}$ we are able to find a term for the distance $dis(\overline{Y_k}, \overline{Y_\infty})$ written only in terms of k. In fact, the distance is $dis(\overline{Y_k}, \overline{Y_\infty}) = O(1/k)$, so $\overline{Y_k} \rightrightarrows \overline{Y_\infty}$. $\square$

Theorem 7.3 *For fixed $k > 0$ and $p > 0$*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k(p)} = 1 - \left(\frac{2k-1}{2k}\right)^{p-1}. \tag{19}$$

Proof. Simple calculation on limits. We are going to combine together formula (13) with the main result given in formula (14). $\square$

The result shows how big asymptotically the size of the fraction of simple tautologies with p premises among all formulas of p premises is. We can see that with p growing this fraction becomes closer and closer to 1. Of course the fraction of all, not only simple, tautologies with p premises is even larger. So the *"density of truth"* within the classes of formulas of p premises can be as big as we wish. For every $\varepsilon > 0$ we can effectively find p such that among formulas with p premises almost all formulas (except a tiny fraction of the size ε) asymptotically are tautologies. This should be contrasted with the results proved in theorems 5.1. It shows that *density of truth* for all $p's$ together is always of the size $O(1/k)$. The result for every p treated separately is very different. Based on theorem 7.3 we may try to estimate the probability for a random long implicational formula to be a tautology by the *"fuzzy"* algorithm bellow.

Given: Implicational formula ϕ from $\mathcal{F}_k^{\{\to\}}$.

Problem: Estimate the chances for ϕ to be a tautology.

Solution: Find p, the number of premises of ϕ which can be done quickly in $log(n)$ time. Then the chances are about $1-((2k-1)/2k)^{p-1}$.

The lack of precision of this *"fuzzy"* algorithm is caused by two reasons. The set of all tautologies is larger than the set of simple tautologies and, moreover, the asymptotic density may be different than the real proportion between the number of simple tautologies and all formulas. We do not have right now the precise estimation of the accuracy of the answer given by the algorithm above.

References

[1] Comtet, L., "Advanced combinatorics. The art of finite and infinite expansions," Revised and enlarged edition, Reidel, Dordrecht. (1974).

[2] Flajolet P., Sedgewick R. "Analitic combinatorics, symbolic combinatorics," unpublished, see `http://algo.inria.fr/flajolet/Publications/books.html`

[3] Moczurad M., Tyszkiewicz J., Zaionc M. *Statistical properties of simple types,* Mathematical Structures in Computer Science, **10** (2000), 575-594.

[4] Robins, H. (1955) *A remark on Stirling formula,* Ammer. Math. Monthly **62** (1955), 26-29.

[5] Szegö, G., "Orthogonal polynomials", Fourth edition, *AMS, Colloquium Publications*, **23**, Providence, 1975.

[6] Wilf H.S., "Generatingfunctionology", Second edition. *Academic Press*, Boston. (1994).

[7] Zaionc, M. *On the asymptotic density of tautologies in logic of implication and negation,* submitted, (see at `http://www.ii.uj.edu.pl/~zaionc/papers.html`

[8] Kostrzycka M., Zaionc, M., *Statistics of intuitionistic versus classical logics,* accepted in Studia Logica.

Acknowledgements: I would like to thank anonymous referee for valuable hints which improve the content and the presentation of our result as well as for some ideas for future research in this topic.
The part of symbolic computations especially calculation of the cartesian distance between distributions presented in the proof of theorem 7.2 has been done using *Mathematica*® [3].

[3] *Mathematica* is a registered trademark of Wolfram Research